The generalization system and method in accordance with the invention solves the above problems and permits similar elements in a web page to be treated in the same manner so that a dynamic web page may be processed using the guide. In particular, an element may occur an arbitrary number of times in the web page without disrupting the automatic processing using the guide. For example, a newspaper home web page may have one or more top newstories. If an extra top newstory is added to the home web page, the guide intended to process the original home page will also automatically process the home page with the extra top newstory.

In more detail, the generalization system and method involves a combination of user input and automatic processing and computation. In this method, the user selects an example of a type of group or atomic that may dynamically change in number and then adjusts the amount of content that is represented by the element. For example, the user may elect to remove certain elements from the new selected content or to move further up or down the XHTML tree to make the content selection larger or smaller. The user then views the selection and either approves the change or provides more input. In a preferred embodiment, the goal of the generalizer is to compute XPath expressions that represent a set of selected nodes in an XHTML page, the number of which might change from page to page or from time to time.

Thus, in accordance with the invention, a system and method for generalizing a set of varying number of atomics and/or groups in a hierarchical document structure (e.g., XHTML or XML) is provided. The method may include identifying an anchor node where the anchor node is defined as the context XHTML node of the XSL template for a particular RML node and identifying an anchor node parent with sibling delimiters where, each item shares the same parent. However, if there are

other items that are identical and also share the same parent, they should not be included. The method further comprises identifying an anchor node sibling where each individual area of generalized structure is not capable of being contained underneath its own unique ancestor node. Typically, this occurs when each of the examples spans several nodes underneath a parent common to all of them. In this case, the anchor node is not a parent of all of the remaining XPath expressions within the template. Instead, the anchor node is a sibling to the first node in each XPath. The method further comprises identifying an anchor node sibling with tangling where due to the way tables are structured in HTML, it is easy for structured areas that are divided into rows and columns to become tangled. With the above methods, the generalizer could easily handle generalization of individual rows or individual columns. However, generalization of tabled data posed a problem because the anchor node computed happened to be shared by multiple examples. This caused the general XPath expressions within the template to match more than one item.

The method further comprises generating an XPath expression that represent a set of selected nodes in an XHTML page, the number of which might change from page to page or from time to time, and generating a generalized XPath expression for a set of atomics and/or groups in an XHTML page.

Brief Description of the Drawings

Figure 1 is a diagram illustrating an embodiment of the generalizer system and method implemented on a typical computer system;

Figure 2A and 2B are diagrams illustrating the generalizer system incorporated into a wireless web page generation system;

Figure 3 illustrates an example of generalization;

Figure 4 illustrates a context node;

Figure 5 illustrates an embodiment of a generalizer method in accordance with the invention;

Figure 6 illustrates more details of the path combiner step of the method shown in Figure 5;

Figure 7 illustrates more details of the node untangler step of the method shown in Figure 5;

Figures 8A - 8C illustrate a first generalizer example for generalizing atomics within a group in accordance with the invention;

Figures 9A - 9C illustrate a second generalizer example for generalizing atomics within a group (multiple groups) in accordance with the invention;

Figures 10A - 10C illustrates more details of the second example of the generalization shown in Figures 9A and 9B;

Figures 11A - 11D illustrate a third generalizer example for generalizing multiple groups in a row-wise manner in accordance with the invention;

Figures 12A - 12C illustrate a fourth generalizer example for generalizing multiple groups in a column-wise manner in accordance with the invention;